NUCLEIC ACID MOLECULE FOR ENHANCING GENE EXPRESSION

Pursuant to 35 U.S.C. §202(c), it is hereby acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Institutes of Health, Grant No. 01CA56110.

FIELD OF THE INVENTION

This invention relates to the field of recombinant DNA technology. More specifically, the invention pertains to DNA constructs and methods for enhancing gene expression to produce clinically beneficial proteins in transformed cells.

BACKGROUND OF THE INVENTION

Several publications are referenced in this application by numerals in parentheses in order to more fully describe the state of the art to which this invention pertains. Full citations for these references are found at the end of the specification. The disclosure of each of these publications is incorporated by reference herein.

Advances in the field of molecular biology over the last two decades have made possible the identification and detailed study of genetically significant regions of specific DNA molecules.

For a gene to be transcribed it is necessary for specific protein factors known as transcription factors to bind particular sites in the regulatory regions of the gene to induce its transcription by the enzyme RNA polymerase. Certain transcription factors, such as TFIID, are constitutively expressed and are required for the assembly of a basal, stable transcription complex.

A promoter is a specific DNA sequence that signals where RNA synthesis should begin. The level of

transcription directed by the basal transcriptional complex bound at the promoter is greatly enhanced by the binding of other constitutively expressed factors to upstream promoter elements which are usually located immediately upstream of the promoter itself. In addition, however, many genes contain other regulatory DNA sequences, which are interdigitated with the upstream promoter elements and which bind transcription factors that only become active in specific cell types or in response to a particular signal. Thus, the presence of such sequences can confer a specific expression pattern on a particular gene.

The transcription of eucaryotic genes can also be regulated by more distant elements known as enhancers. Enhancers contain binding sites for the same constitutively expressed or tissue-specific regulatory factors which bind immediately upstream of the promoter but often contain multiple copies of the binding site or sites for many different factors. Although the enhancer cannot drive transcription itself, it can enhance the activity of the promoter by several orders of magnitude. Such enhancement may occur in all cell types if the enhancer contains binding sites for constitutively expressed transcription factors or may occur only in specific tissue or in response to a specific signal if the enhancer contains binding sites for factors which are involved in gene regulation.

Although most constitutively expressed or regulated transcription factors activate the transcription of specific genes, it is also possible for transcription to be specifically inhibited by the action of transcription factors. One mechanism by which a factor can inhibit gene expression is by preventing the binding to DNA of another activating factor. Another mechanism involves the negative factor binding to the already DNA-bound activating factor resulting in a masking of the activation domain. Finally, some negative

factors can inhibit gene expression directly. Regardless of the molecular mechanism, the binding of such factors to DNA results in the repression of transcription. In situations requiring maximal gene expression of a protein of interest, identification of these negative sites in relevant promoter regions is highly desirable.

DNA vectors (i.e., plasmids, viruses) can be modified to include a gene for a foreign protein. Some of these recombinant vectors contain a promoter/enhancer necessary for the expression of the gene upstream from what is known as a cloning site. (i.e., a rare or unique restriction enzyme site where a foreign gene of interest can be inserted). The recombinant expression vectors can then be transfected into cells or tissues where the foreign gene is to be expressed.

Such expression vectors are of great value for the purposes of creating transgenic animals and for use in gene therapy. Current strategies involving gene therapy seek to first identify the defective gene, and then to supplement the defective tissues with the functional gene. Transgenic animals and gene therapy are only two examples of the wide range of uses for expression vectors in biology and medicine. Expression vectors also have many *in vitro* uses, such as permitting production of large amounts of a protein of interest.

Viral recombinant promoter elements are currently the most widely used in the construction of recombinant vectors. Typically, they are derived from pathogenic viruses such as Epstein Barr Virus (EBV), Rous Sarcoma Virus (RSV), Simian Virus 40 and Human Cytomegalovirus (CMV). While generally effective, in some cases viral promoters are "turned off" or repressed after a time when in a eucaryotic host.

Accordingly, there is a need for the identification of sequences required for the binding of the above-mentioned putative transcriptional repressors. Identification and elimination of these repressor binding

sites in viral or cellular promoter sequences will facilitate the development of recombinant vectors for use in vitro and in gene therapy that demonstrate enhanced gene expression and measurable increases in the production of the encoded, clinically beneficial proteins.

SUMMARY OF THE INVENTION

The present invention provides DNA constructs, expression vectors and methods for their use in augmenting gene expression in cultured cells and in animals, including humans. The DNA constructs and vectors of the invention are used beneficially in gene therapy and for the production of DNA vaccines.

In accordance with the present invention, an isolated DNA construct is provided, which comprises at least one mutated binding site for a Gfi-1 transcription repressor. The mutated binding site comprises a mutation which hinders or prevents binding of Gfi-1 to the site. The mutated binding site preferably is disposed within an expression regulatory segment, such as a promoter or enhancer, and most preferably within a mammalian cellular promoter or a viral promoter, such as a human cytomegalovirus (CMV) promoter.

In a preferred embodiment, the aforementioned Gfi-1 binding site, prior to mutation, is at least 65% (more preferably at least 79%, and most preferably at least 98%) homologous with a sequence comprising N_1AAATCACN_2GCA (Sequence I.D. No. 1), wherein N_1 and N_2 are any nucleotide, but preferably N_1 is T and N_2 is A or T (Sequence I.D. No. 2 is the preferred sequence, TAAATCAC(T/A)GCA). In a particularly preferred embodiment, the mutation is in that portion of the binding site having the sequence AATC.

According to another aspect of the invention, the DNA construct disposed within an expression regulatory segment is operably linked to a coding

segment. The coding segment preferably encodes gene products such as cytokines, interleukins, interferons, growth factors and signalling proteins involved in regulation of cell growth (oncogenes).

According to another aspect of the invention, an expression regulatory DNA segment is provided which comprises at least one copy of a sequence at least 65% (more preferably at least 79% and most preferably at least 98%) homologous with the sequence N1A-R-CN2AGCA (Sequence I.D. No. 3). Here, N_1 and N_2 are any nucleotide, with preferred designations as described above, and R is a tetranucleotide selected from the group consisting of:

N_3ATC, AN_4TC, AAN_5C, $AATN_6$

N_3N_4TC, N_3AN_5C, N_3ATN_6, AN_4N_5C, AN_4TN_6, AAN_5N_6

$N_3N_4N_5C$, $N_3N_4TN_6$, $N_3AN_5N_6$, $AN_4N_5N_6$, and $N_3N_4N_5N_6$,

wherein N_3 is G, C or T, or is absent, or is an oligonucleotide of two or more nucleotides; N_4 is G, C or T, or is absent, or is an oligonucleotide of two or more nucleotides; N_5 is A, G or C, or is absent, or is an oligonucleotide of two or more nucleotides; and N_6 is A, C or C, or is absent, or is an oligonucleotide of two or more nucleotides. In a preferred embodiment, the expression regulatory segment is a promoter, most preferably a mammalian cellular promoter or a viral promotor, such as a CMV promoter.

According to another aspect of the invention, the aforementioned expression regulatory sequence is provided in an expression vector, which also contains an operatively positioned insertion site for insertion of a coding segment. In a preferred embodiment the coding segment encodes gene products such as cytokines, interleukins, interferons, growth factors or signalling proteins involved in regulation of cell growth (oncogenes).

According to another aspect of the invention, a mutant promoter is provided, which comprises two mutated

Gfi-1 binding sites. This promoter is derived from the CMV-MIE wildtype promoter (Sequence I.D. No. 12), and comprises sequences such as those exemplified by, but not limited to, Sequence I.D. No. 13 or Sequence I.D. No. 14, both of which contain mutations in the two naturally occurring Gfi-1 binding sites found in the wildtype promoter. This promoter is preferably contained within an expression vector.

According to yet another aspect of the present invention, methods are provided for improving the expression of genes regulated by expression regulatory sequences that contain binding sites for the Gfi-1 transcription repressor. The methods comprise altering the sequence of the Gfi-1 binding sites in the regulatory segments, so as to hinder or prevent binding of Gfi-1 to those sites, thereby improving expression of the genes.

The present invention facilitates maximized exogenous gene expression by expression regulatory segments (such as the CMV-MIE promoter) containing recombinant vectors in cells. Such vectors can be used to direct expression of any gene.

Vectors containing a foreign gene of interest under the regulation of the promoter are administered to cells of a living organism under conditions whereby the vector enters cells and expresses the protein encoded by the gene of interest. Such gene expression can occur without repression by Gfi-1 in the cells, since the promoter contains mutated Gfi-1 binding sites to which the transcription repressor cannot bind.

According to another aspect of the present invention, a method is provided for treating a pathological condition related to the expression of an aberrant gene. The method comprises administering to a patient having such a pathological condition a pharmaceutical preparation comprising a vector of the invention, as described above, capable of entering a cell expressing the aberrant gene. Expression of the normal

gene results in production of a functional gene product, thereby alleviating the pathological condition.

According to another aspect of this invention, a pharmaceutical preparation is provided for treating a pathological condition related to the aberrant gene expression. This pharmaceutical preparation comprises, in a biologically compatible medium, a vector of the invention, as described above, encoding a gene of interest capable of entering a cell, and expressing the protein encoded thereby. Incorporation of the DNA constructs of the invention into suitable recombinant vectors for delivery via liposomes is also contemplated to be within the scope of this invention.

The DNA expression vector encoding the gene of interest is synthesized so as to be capable of crossing a biological membrane in order to enter cells and thereafter express the protein encoded by the gene of interest. A biologically compatible medium is preferably formulated to enhance the lipophilicity and membrane-permeability of the expression vector.

In yet another preferred embodiment, the DNA constructs and the vectors of the invention may also be used alone or in combination with chemotherapeutic drugs to treat bone marrow or peripheral stem cell grafts of cells. They may also be used to deliver DNA vaccines.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a segment of DNA containing the deduced Gfi-1 binding site (Sequence I.D. No. 4), deduced by PCR amplification of a random library. The AATC motif is underlined.

Figure 2 is a composite of data illustrating the significance of specific bases on Gfi-1/DNA binding. Figure 2A is an autoradiogram of DNAse I footprint analysis using recombinant GST/Gfi-1 protein and near consensus oligonucleotide. Figure 2B is an autoradiogram showing the results of DMS methylation interference.

This identifies the bases in both DNA strands that come in contact with the GST-Gfi-1 fusion protein. Figure 2C (Sequence I.D. No. 5) shows the combined data from figures 2A and 2B. The solid line indicates the extent of the DNA footprint. The circles above identify individual bases that are in contact with the protein. Figure 2D1 shows the mutant binding sites of the oligonucleotides. R21 is Sequence I.D. No. 6; R21A is Sequence I.D. No. 7; R21B is Sequence I.D. No. 8; R21C is Sequence I.D. No. 9; R21D is Sequence I.D. No. 10; and R21E is Sequence I.D. No. 11. Figure 2D2 is an autoradiogram of EMSA performed on the oligonucleotides of Figure 2D1. Figure 2D3 is a histogram quantitating the results in Figure 2D2. Figure 2E illustrates that Gfi-1 from transfected COS-1 nuclear extracts binds oligonucleotides encoding the R21 binding site but not to the mutant site, i.e., R21A.

Figure 3A shows the nucleic acid sequence of the HCMV-MIE promoter (Sequence I.D. No. 12). The two putative Gfi-1 binding sites are boxed. The four 18 bp repeats are underlined. CAAT and TATA boxes are shown with a double underline. The major transcription start site is indicated by +1. Figure 3B shows the nucleotide sequence of the two potential Gfi-1 binding sites (as shown in Sequence I.D. No. 12) compared to the Gfi-1 binding consensus sequence (Sequence I.D. No. 1). Figure 3C shows the nucleotide changes in the mutated Gfi-1 binding sites in the CMV-MIE promoter. Mutant A is Sequence I.D. No. 13 and Mutant B is Sequence I.D. No. 14. Figure 3D is a graph showing the relative CAT activity in extracts of NIH-3T3 cells transfected with wild type and mutant HCMV-MIE promoter/CAT reporter constructs in the presence or absence of Gfi-1.

Figure 4 shows a schematic representation of the contacts between Gfi-1 amino acids and the nucleotides of the binding site.

DETAILED DESCRIPTION OF THE INVENTION

A. Definitions

The nucleic acids of the invention are sometimes referred to herein as "isolated nucleic acids". This term, when applied to DNA, is intended to signify a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' to 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acids" of the invention may comprise a DNA molecule inserted into a vector, such as a plasmid or a virus vector, or integrated into the genomic DNA of an organism. With respect to RNA molecules, the term "isolated nucleic acids" primarily refers to RNA molecules encoded by isolated DNA molecules as defined above, or produced by synthetic methods.

The terms "transcription control element" and "expression regulatory segment" are used interchangeably herein, and refer to an isolated DNA segment that, under specified conditions, possesses a transcription-regulating activity with respect to the expression of a coding segment that encodes a gene product. Expression regulatory segments include promoters, enhancers, internal elements and 3' regulatory segments.

The term "transcription factor" refers to protein factors that interact with DNA to enhance or repress transcription.

The term "repressor" refers to a transcription factor that inhibits gene expression upon interaction with DNA.

The term "transcription unit" refers to a nucleic acid molecule comprising one or more sequences (referred to herein as a "coding segment") that encodes a gene product (usually a protein) and is operably linked to a promoter or other expression regulatory sequences (as defined above) necessary for expression of the coding

sequence. The term "gene" is often used interchangeably with the term "transcription unit".

The term "operably linked" or "operably positioned" means that the regulatory segments necessary for expression of the coding segment are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to effect expression of the coding sequence.

The term "gene therapy" refers to methods for augmenting levels of an expressed gene or replacing a defective gene via delivery of the vectors of the invention containing the gene of interest to specified cell types or tissues.

When used herein in describing DNA fragment lengths or other experimental results, the term "approximately" means within a margin of commonly acceptable error for the determination being made, using standard methods (e.g., agarose gel electrophoresis and comparison with DNA molecules of known size to determine DNA fragment size or relative position).

When referring to specific nucleic acid sequences set forth herein, the term "substantially the same as" means taking into account minor variations or substitutions that arise for a number of reasons, but do not alter the overall characteristics of the DNA molecule defined by the sequence. For example, homologous regions isolated from different species, sub-species or strains of an organism may possess sequence polymorphisms that render those sequences substantially the same as, but not identical to, the sequences set forth herein. With respect to the nucleic acids, substantial similarity is generally determined by sequence homology. Accordingly, the present invention is intended to encompass all sequences that are "substantially the same" as the sequences exemplified herein, within the confines of appropriate levels of sequence homology. Specifically, the invention is intended to encompass isolated nucleic

acid molecules having at least about 65% (preferably at least 80%, and most preferably at least 90%) sequence homology with the exemplified seqeuences, unless otherwise specified.

B. Description

The *growth factor independence*-1, (Gfi-1) gene encodes a zinc finger protein with six C_2H_2 type C-terminal zinc finger motifs. Gfi-1 is a nuclear protein which binds DNA in a sequence specific manner, and functions as a transcriptional repressor. DNA binding is mediated by three of the six Gfi-1 zinc fingers. Gfi-1 binding to DNA, *in vitro*, generates a 21 bp footprint which extends over the 12 bp binding site, N_1AAATCACN_2GCA (Sequence I.D.No. 1, wherein N_1 preferably is T and N_2 preferably is A or T). DMS methylation interference analysis identified nine bases that closely interact with the protein. Point mutations of the binding site diminished or abrogated Gfi-1 binding and confirmed the specificity of the site. Although methylation interference suggested that Gfi-1 binding depends on protein DNA contacts both within and outside the AATC core, electrophoretic mobility shift assays using oligonucleotides of the wild type or mutant binding sites showed that only residues within the AATC core are critical.

Among the genes whose promoters contain Gfi-1 binding sites are genes encoding various cytokines and other regulators of cellular proliferation and differentiation (Table II in the Examples below). The induction of Gfi-1 during T cell activation, and its upregulation by provirus insertion during oncogenesis are expected to repress the expression of such genes. Gfi-1, therefore, may contribute to T cell activation, a process characterized by sequential waves of expression of cytokine and other regulatory genes (17), by regulating their timely repression. Since the expression of such

genes may induce growth arrest, differentiation, and apoptosis in certain cell types, their repression by Gfi-1 may also contribute to the process of oncogenesis.

The DNA constructs of the invention comprise mutated Gfi-1 binding sites as described above, which are modified so that they are no longer recognized by the repressor for binding, and binding of the repressor is thereby eliminated. These mutated binding sites are preferably incorporated into vectors, such as plasmid or viral vectors, such that they are operably linked to, and control the expression of, coding sequences contained within such vectors. Methods for the synthesis of such DNA constructs and the assembly of vectors for expression of coding sequences are well known in the art. They include oligonucleotide or DNA synthesis, restriction enzyme digestion, annealing and ligation of nucleic acid fragments, and other common methods as described generally, for example, by Sambrook et al.

As discussed above, Gfi-1 binding sites have been identified in several cellular and viral promoter sequences, and use of such promoters with mutated binding sites is preferred for practice of the invention. However, to the extent that Gfi-1 binding sites are also present in other exrpression regulatory sequences, e.g., enhancers, internal sequences or 3' sequences, mutants of these regulatory segments that do not bind Gfi-1 are also contemplated for use in the present invention. Elucidation of the Gfi-1 consensus binding sequence in accordance with the present invention enables identification of such binding sites wherever they may occur in a DNA molecule, regardless of location.

One of the promoters containing Gfi-1 binding sites is the HCMV-MIE promoter, which regulates the expression of genes required for the initiation of HCMV infection and probably the reactivation of HCMV from latency (18, 19). This promoter contains two sites with 79% and 80% homology to the Gfi-1 binding site consensus.

HCMV-MIE promoter/CAT reporter constructs were downregulated when cotransfected with Gfi-1 in NIH 3T3 fibroblasts. Moreover, the Gfi-1 mediated repression of the MIE promoter was abrogated when the Gfi-1 binding sites were mutated.

This abrogation of Gfi-1 mediated repression can be used to advantage to enhance the expression of genes driven by the CMV-MIE promoter. Gene therapy approaches utilitizing vectors containing the promoter operably linked to the coding region of a protein of interest followed by introduction into cells is expected to result in a measurable increase in protein expression.

The CMV-MIE promoter has been used frequently to drive expression of exogenous genes for *in vitro* use, for construction of DNA vaccines and for other forms of gene therapy. The current invention exemplifies the mutational analysis of DNA sequences in the CMV-MIE promoter. However, as mentioned, Gfi-1 binding sites are detectable in a variety of viral and cellular expression regulatory sequences, and mutation of such sites to hinder or prevent the binding of the Gfi-1 transcription repressor also results in increased expression of such genes.

The following examples, relating to identification of the Gfi-1 consensus binding sequence and mutation of these binding sites in the HCMV promoter, set forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for the purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning and other recombinant DNA procedures are used, such as those set forth in Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.").

MATERIALS AND METHODS FOR EXAMPLES I - III

A. Preparation of Gfi-1 Bacterial Fusion Proteins:

Expression constructs of Gfi-1 in the pGEX vector (Pharmacia) were used to express glutathione-S-transferase (GST)/Gfi-1 fusion proteins in *Escherichia coli*. Different constructs were designed to express the entire Gfi-1 protein, the zinc finger domain, and the zinc finger domain carrying deletions of individual zinc fingers. Constructs with deletion of individual zinc fingers were generated by overlap extension PCR (1). *E. coli* transformed and selected in ampicillin were grown to log phase and induced with 1 mM isopropyl 13-D-thiogalactopyranoside (IPTG) (Sigma). Three hours following induction, the bacteria were centrifuged and sonicated in cold phosphate buffered saline (PBS). The lysates were then clarified by centrifugation at 15,000 g for 10 minutes at 4°C and were mixed with glutathione-linked agarose beads (Sigma) (1/1000th the volume of the induced bacterial culture). Following a 10 minutes incubation, the beads were washed three times in cold PBS. If the fusion protein was to be eluted, the pelleted beads were washed with 10 mM reduced glutathione (Sigma) in 50 mM Tris-HCl pH 8.0 (2). The liquid was collected and concentrated in a Centriprep-30 column (Amicon). Protease inhibitors, Aminoethylbenzene sulfonylflouride (AEBSF) (Calbiochem) and Aprotinin (Sigma)) were added prior to freezing.

B. Gfi-1 DNA Binding and Random Oligonucleotide Selection:

A 54-base single-stranded DNA oligonucleotide was synthesized to contain a central region of 18 random bases flanked by 18-base regions with defined sequences. One overlapping, complementary oligonucleotide was annealed and extended by AmpliTaq polymerase (Perkin Elmer) to yield a mixture of double-stranded DNA fragments containing >500 copies of each possible

permutation. This mixture was incubated with GST/Gfi-1 fusion protein bound to beads, prepared as described above, in Binding Buffer (0.2 mg/ml poly dI-dC (Sigma), 0.2 mg/ml acetylated BSA (NEBiolabs), 25 mM Hepes pH 7.5, 100 mM KCl, 0.1 mM $ZnSO_4$, 10 mM $MgCl_2$, 0.1% NP-40, 1 mM DTT, 5% glycerol) for 30 minutes. The beads were centrifuged and washed four times with Binding Buffer, then boiled for 5 minutes in H_2O. The oligonucleotide mixture eluted from the beads by boiling was used for PCR amplification using the same amplifying oligonucleotides (3). After 4 rounds of selection/amplification, the PCR products were digested with *Eco*RI and *Hind*III and cloned into pBluescript (Stratagene). Alkaline-lysis prepared plasmid DNA from individual *E. coli* transformants (4) were sequenced (USB Sequenase 2.0 kit).

C. Electro-mobility Shift Assays (EMSA), DNAse I Footprinting and DMS Methylation Interference:

Double-stranded DNA fragments of selected Gfi-1 binding sites were produced by PCR amplification of plasmid templates using oligonucleotides end labeled with [γ32P]ATP and T4 Polynucleotide Kinase (New England Biolabs). PCR products and Gfi-1 protein were incubated in EMSA Binding Buffer (4 mM Tris-HCl pH 7.5, 80 mM NaCl, 0.5 mM $ZnSO_4$, 1 mM EDTA, 0.5 mM DTT, 1 μg poly dI-dC, 1 μg poly dA-dT (Sigma), 5% glycerol) at 22°C for 30 minutes. Samples were electrophoresed for approximately 3 hours at 100 volts on pre-run, non-denaturing polyacrylamide gels (6% (29:1) Acryl:Bisacrylamide, 0.25xTBE, 5% glycerol) (5). To determine the relative importance of specific bases for Gfi-1 binding, EMSAs were performed with 24 base pair long double stranded mutant oligonucleotides. These oligonucleotides were synthesized *in vitro* and end-labeled using [γ-32P]ATP and T4 polynucleotide kinase.

DNase I footprinting was carried out using a modified published technique (6). PCR products of

selected Gfi-1 binding sites, 5' labeled with ^{32}P, and recombinant Gfi-1 protein were mixed in 100 µl DNase I Buffer (10 mM Tris-HCl pH 7.8, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5 mM $ZnSO_4$, 0.5 mM EDTA, 0.1 mM DTT, 0.1 mg/ml acetylated BSA, 2 µg poly dI-dC, 2% polyvinyl alcohol, MW 30-70 kDa (Sigma), 5% glycerol) and incubated at 4°C for 30 minutes. The DNA was then digested for 2 minutes at 4°C by adding approximately 2 µg DNase I (Sigma). The reaction was stopped by adding a 1/1 mixture of phenol:chloroform. The partially digested DNA was ethanol precipitated in the presence of 10 µg glycogen (Boeringer Mannheim) and electrophoresed in a standard 8% polyacrylamide, 8 M urea gel. The fractionated DNA fragments were visualized by autoradiography and phospho-imaging (Fuji) for quantitation.

DMS methylation interference assays were carried out using a modified standard protocol (7). PCR-amplified Gfi-1 binding site oligonucleotides, 5' labeled with ^{32}P, were mixed with 1 µl dimethylsulfate (DMS) in 100 µl DMS Buffer (50 mM Sodium Cacodylate, 1 mM EDTA, 4 µg poly dI-dC) at 22°C for 2 minutes. The methylation reaction was stopped with 25 µl DMS Stop Buffer (0.86 M Tris-HCl pH 7.5, 1.5 M Sodium acetate, 1 M 2-mercaptoethanol). The DNA was then ethanol precipitated, washed twice, and resuspended in Tris-EDTA. The partially methylated DNA was incubated with and without Gfi-1 protein and was subsequently electrophoresed in a nondenaturing polyacrylamide gel. Input DNA (not incubated with Gfi-1) and shifted DNA (Gfi-1 bound) were extracted from crushed gel fragments by incubation in Acrylamide Elution Buffer (0.5 M NH_4OAc, 10 mM MgOAc, 1 mM EDTA, 0.1% sodium docedecyl sulfate (SDS)) for 3 hours at 37°C. The eluted DNA was ethanol precipitated with 10 µg glycogen. The DNA pellet was resuspended in 100 µl 1 M piperidine, heated to 90°C for 30 minutes, ethanol-precipitated and washed three times,

and electrophoresed in standard 8% polyacrylamide, 8 M urea sequencing gels.

D. Transient Transfection of COS-1 Cells and Gfi-1 EMSA with COS-1 Nuclear Extracts:

COS-I cells (8) were transiently transfected with pCMV5/Gfi-1 or pCMV5 (9) using DEAE Dextran as described for NIH 3T3 cells (10). Forty-eight hours later the transfected cells were washed, scraped and pelleted at 4°C into microcentrifuge tubes. Cell pellets were then resuspended in 10 volumes buffer A (10 mM Hepes-KOH pH 7.9 at 4°C, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol (DTT), 0.2 mM AEBSF) for 10 minutes at 4°C, they were then vortexed for 10 seconds and centrifuged in a microfuge for 10 seconds. The pelleted nuclei were resuspended in 2 volumes buffer C (20 mM Hepes-KOH pH 7.9 at 4°C, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM AEBSF, 25% glycerol) for 20 minutes at 4°C. The resulting lysates were centrifuged at 15,000 g at 4°C for 5 minutes and then were quick frozen and stored in aliquots at -70°C (11). DNA binding for EMSAs was carried out in a final volume of 20 μl using 5 μl of lysate, 1 μg poly dI-dC, 1 μg poly dA-dT, and ^{32}P-labeled DNA generated by PCR.

E. Human Cytomegalovirus (HCMV) Major Immediate Early (MIE) Promoter and Chloramphenicol Acetyl Transferase (CAT) Assay:

The MIE promoter of HCMV (12) contains two putative Gfi-1 binding sites, suggesting that it is regulated by Gfi- 1. To test this hypothesis, both sites (13) in the HCMV-MIE promoter were mutated in the pCMV5 expression vector. The CAT gene was inserted in both wild type and mutant vectors. The wild type and mutant promoter constructs were transfected into NIH 3T3 cells using Lipofectamine (Gibco BRL) according to the procedures recommended by the manufacturer. Forty-eight

hours later, the cells were washed with PBS, and collected in 0.25 M Tris-HCl pH 7.8. Cell pellets were frozen in an ethanol/dry ice bath, thawed at 37°C for 4 minutes, and vortexed. The lysates generated by repeating this process four times were clarified by centrifugation at 15,000 g at 4°C for 5 minutes. Equivalent amounts of lysates were mixed at 22°C with coenzyme A (CoA) Reaction Mix (80 mM Tris-HCl pH 7.8, 360 μg/μl aqueous chloramphenicol, 67 μM acetyl CoA, 3 μCi [^{3}H]acetyl CoA (Amersham)) in a scintillation vial (35). Scintillation fluid (Econofluor, Dupont) was gently overlayed to fill the vial. ^{3}H counts were measured at successive time points in a scintillation counter (Beckman LS 6000IC). Data presented in Figure 3D represent the average of six independent transfections. Similar data were obtained in three separate experiments.

EXAMPLE I

IDENTIFICATION OF THE Gfi-1 CONSENSUS BINDING SITE

Gfi-1 is a 55 kD nuclear protein that binds DNA in a sequence specific manner. The Gfi-1 binding site TAAATCAC(A/T)GCA was defined via random oligonucleotide selection utilizing bacterially expressed GST/Gfi-1 fusion protein. Binding to this site was confirmed by electrophoretic mobility shift assays (EMSA) and DNase I footprinting. DMS methylation interference assay and EMSA with mutant oligonucleotides defined the relative importance of specific bases in the consensus binding site. Potential Gfi-1 binding sites were detected in a large number of eukaryotic promoters/enhancers, including the enhancers of several protooncogenes and cytokine genes and the enhancer of the HCMV-MIE which contains two such sites. HCMV-MIE promoter/CAT reporter constructs, transfected into NIH 3T3 fibroblasts, were repressed by Gfi-1, and the repression was abrogated by mutation of critical residues in the two Gfi-1 binding sites. These results suggest that Gfi-1 may play a role in HCMV

biology and may contribute to oncogenesis and T cell activation by repressing the expression of genes that inhibit these processes.

A prerequisite for the predicted role of Gfi-1 in transcription is that the Gfi-1 protein is localized in the nucleus. To determine the subcellular localization of Gfi-1 NIH 3T3 cells were transfected with a pCMV5/Gfi-1 expression construct tagged at its N-terminus with a twelve amino acid hemagglutinin (HA) epitope tag. Western blots of total cell lysates from untransfected and transfected cells probed with the HA tag monoclonal antibody 12CA5 revealed that Gfi-1 encodes a 55 kD protein (data not shown). Immunofluorescence staining of transfected and untransfected cells with the HA epitope tag antibody demonstrated that Gfi-1 is a nuclear protein (data not shown).

To determine the putative Gfi-1 binding site, a random oligonucleotide binding/selection strategy was used. The library of oligonucleotides selected by four rounds of Gfi-1 binding was cloned in pBluescript and 96 independent clones were sequenced. Of the sequenced clones, 54 (56%) contained one AATC motif, 35 (36%) contained two AATC motifs and 7 (7%) contained none. Aligning the sequences of the 54 clones that contained only one AATC to this motif (Fig. 2) enabled the determination of the Gfi-1 DNA binding consensus as shown in Figure 1, i.e., N_1AAATCACN_2GCA (Sequence I.D. No. 1, wherein N_1 preferably is T and N_2 preferably is A or T).

To confirm the consensus binding site, the Gfi-1 binding/oligonucleotide selection experiment was repeated starting with oligonucleotides containing an AATC motif flanked by 15 bp random DNA sequences and 18 bp regions with defined sequence. The library of selected oligonucleotides after three rounds of Gfi-1 binding was cloned in pBluescript. Sequencing of 24 clones confirmed the consensus binding site TAAATCACNGCA.

The homology of selected Gfi-1 binding sites to the consensus ranged from 65% to 98%. This finding tended to indicate a correlation between the relative binding affinity of Gfi-1 to these sites and their homology with the consensus sequence. In order to determine whether this is so, six selected oligonucleotides whose homology to the consensus binding site ranged from 65% to 98% were used to carry out Gfi-1 binding and electrophoretic mobility shift assays. The Gfi-1 protein used in these experiments was expressed as a glutathione-S-transferase (GST) fusion protein in *E. coli*. The results showed that indeed the relative binding affinity of Gfi-1 to individual selected sites is proportional to their homology to the consensus (data not shown).

EXAMPLE II

DNA FOOTPRINT ANALYSIS OF Gfi-1 BINDING SITE

To define the DNA region to which Gfi-1 binds, oligonucleotides representing 4 selected Gfi-1 binding sites with 72, 85, 97 and 98% homology to the consensus were incubated with bacterially expressed GST/Gfi-1 fusion protein. The DNA/protein complexes were digested partially with DNase I and electrophoresed in polyacrylamide/urea sequencing gels. The protected DNA region in each oligonucleotide was assessed by autoradiography. Footprints were obtained only when using oligonucleotides with 97 and 98% homology to the consensus. The results showed a 21 base pair footprint extending 8 bases 5' and 9 bases 3' of the AATC motif in both cases as shown in Fig. 2A.

To define the bases in the Gfi-1 binding site that are in contact with the protein, DMS methylation interference assays were performed using PCR amplified Gfi-1 binding site oligonucleotides and bacterially expressed GST/Gfi-1 fusion protein. The results showed that nine individual bases, when methylated, interfered

with Gfi-1 binding. See Fig. 2B. Specifically, the methylation of guanines on the complementary strand at positions 6 and 8 greatly impaired protein DNA binding, while methylation of adenines within and around the AATC core and guanines distal to the core had a moderate effect. The footprint and methylation interference data are summarized in Fig. 2C.

Mutations in the Gfi-1 binding site abolish Gfi-1 binding. To confirm the significance of individual bases for Gfi-1 binding, a wild type oligonucleotide (R21) and five mutants (R21 A-E) were synthesized. The single base changes in each mutant are underlined. See Table I.

TABLE I

	123456789101112	
R21	CACCACATAAATCACTGCCTATCC	(Seq. I.D. No. 6)
R21A	CACCACATAGATCACTGCCTATCC	(Seq. I.D. No. 7)
R21B	CACCACATAACTCACTGCCTATCC	(Seq. I.D. No. 8)
R21C	CACCACATAAATAACTGCCTATCC	(Seq. I.D. No. 9)
R21D	CACCACATAAATCAATGCCTATCC	(Seq. I.D. No. 10)
R21E	CACCACATAAATCACTTCCTATCC	(Seq. I.D. No. 11)

Using electrophoresis mobility shift assay with recombinant Gfi-1 as the indicator for Gfi-1 binding to these oligonucleotides, the results showed that mutations within the AATC motif greatly decreased or abolished binding, while mutations outside this motif had little effect. See Figs. 2D2 and 2D3. The R21A mutant was still able to bind Gfi-1 although at 5% the level of R21. In agreement with the methylation interference data, a mutation at position 3 prevents binding less efficiently than a mutation at position 4. However, the effect of other mutations could not have been predicted by the methylation interference results alone. Thus, a mutation

at position 6 (R21C) completely ablated binding while a mutation at position 8 (R21D) had little effect despite the fact that the residues occupying both positions appeared to be equally important by methylation interference.

EMSA experiments were also carried out using the wild type R21 and the R21A mutant oligonucleotide and Gfi-1 protein derived from nuclear extracts of transfected COS-1 cells. The results showed that Gfi-1 expressed in COS- 1 cells binds the wild type but not the R21 A mutant oligonucleotide (94% and 5% shifted, respectively) which correlates with the EMSA results using bacterially-expressed Gfi-1. See Fig. 2E. Gfi-1 protein expression in transfected but not in nontransfected cells was documented by Western blotting (data not shown).

Many eukaryotic promoters contain potential Gfi-1 binding sites. A computer database search with the Gfi-1 binding consensus revealed potential Gfi-1 binding sites in the promoter region of many genes (Eukaryotic Promoter Database (14)). Table II presents a list of mammalian DNA polymerase II promoters containing potential binding sites whose sequence is ≥79% identical to that of the consensus. The list includes several proto-oncogenes and cytokine genes. Subsequently we examined whether the Gfi-1 binding site is consistently near, or overlaps with the binding site(s) of other factor(s) (Transcription Factor Database (15)). Such an association could be interpreted to suggest that Gfi-1 binds DNA in concert, or competes with other factor(s) for DNA binding. The results of this analysis revealed that approximately 20% of the putative Gfi-1 binding sites contain GAAATC, the binding site for the histone H4 gene regulatory factor H4TF-I (16).

Table II

Identification of Promoters with Potential Gfi-1 Binding Sites

Promoter		% of Consensus	Gfi-1 consensus TAAATCAC$_{A,T}$GCA	Promoter		% of Consensus	Gfi-1 consensus TAAATCAC$_{A,T}$GCA
IL-1α	Human	80	CAAATCAATAAC	TNF-α	Human	85	CAAATCCCCGCC
IL-1β	Human	86	TAAATCTGTGTG			80	CAAATCAGTCAG
	Mouse	80	GAAATCAGTTAA		Mouse	82	CTAATCATTGTC
IL-4	Human	87	GAAATCAGACCA		Rabbit	86	GAAATCAGAGGG
	Mouse	87	GAAATCAGTTAA			81	CAAATCCGGGTC
IL-5	Human	89	TCAATCACTGTC		Hamster	86	GAAATCAGAGAG
		85	AAAATCCCTGTT	c-mos	Mouse	90	TAAATCACTCCC
		82	AAAATCAGAAAA	c-abl	Mouse	89	TTAATCACAGTC
IL-6	Human	85	TAAATCTTTGTT	c-erb82	Human	88	GGAATCACAGGA
IFNα	Human	86	CAAATCTGTGTT	c-myc	Human	90	TAAATCATCGCA
		84	AAAATCTAAGTT	N-myc	Human	86	AAAATCAGGGGA
	Mouse	91	TAAATCAAAGTT	c-N-ras	Human	85	GAAATCAGACCC
IFNγ	Human	79	GAAATCAGTAGT			81	AAAATCAGTAAA
IGF II	Rat	88	AAAATCTGAGCT		Mouse	84	GAAATCAGGCCA
		87	CAAATCAGACCC			81	AAAATCAGTAAA
		84	CAAATCAGACAA	CD8	Mouse	90	CAAATCTCAGTT
		80	AAAATCTTAGGC	Thy-a	Mouse	88	CCAATCACAGGA
		80	TAAATCCTGGGT	Histone H1A	Human	93	AAAATCAAAGCA
	Human	86	TTAATCACGGTT	LTR	HIV	82	CCAATCAGGGAA
		84	CAAATCCGAGTT	MIE	HCMV	80	AAAATCAACGGG
CSF-1	Human	89	CAAATCTTAGCA			79	GAAATCCCCGTG
		79	GAAATCACCCTG	IEgpUS3	HCMV	87	GAAATCACCGTG
	Mouse	89	CAAATCTTAGCA			87	GAAATCCCAGTA
		79	GAAATCACCCTG	early 2.2kb	HCMV	83	CTAATCACGGAC
G-CSF	Human	79	TAAATCCTGGGA	early 2.7kb	HCMV	84	AAAATCAGTCCG
	Mouse	79	TAAATCCTGGGA	UL36	HCMV	80	GAAATCGCGGGC
c-sis	Rabbit	84	GAAATCAGGCCA	pp65	HCMV	81	CAAATCCACGCT
TNF β	Human	83	CAAATCATACTT			79	AAAATCGGTGGT
	Rabbit	92	CAAATCAGGGCT				

This list includes selected mammalian promoters retrieved by computer search from the Eukaryotic Promoter Database with the Gfi-1 consensus TAAATCAC(A/T)GCA. All listed potential binding sites have scored $\geq$ 79% homology to the consensus. This cutoff represents the value for the HCMV MIE promoter binding sites. (Fig.3).

EXAMPLE III

Gfi-1 REPRESSES THE MIE PROMOTER OF HCMV

Gfi-1 is a transcriptional repressor that downregulates the expression of the MIE promoter of HCMV. To assess the transcriptional regulatory function of Gfi-1 in promoters with Gfi-1 binding sites, a pCMV5/β-galactosidase expression construct was used to monitor transfection efficiency. Surprisingly, the expression of β galactosidase in pCMV5/β-galactosidase transfected NIH 3T3 cells was downregulated in a Gfi-1 dependent manner (data not shown). To determine the specificity of this effect the MIE promoter of HCMV in the pCMV5 vector was examined for the presence of Gfi-1 binding sites. This search revealed two putative binding sites localized within two of the four 18 base pair repeats of this promoter (12). The two sites were 79% and 80% homologous to the Gfi-1 binding site consensus and they were placed at the base and the apex of proposed stem loop structures that may play a role in promoter function (12). The generation of the Gfi-1 binding sites within the two 18 bp repeats was due to a divergence of these repeats from the repeat consensus. See Figs. 3A and 3B.

To confirm and quantitate the Gfi-1 mediated repression of this promoter, HCMV-MIE promoter/CAT reporter constructs were cotransfected transiently with a Gfi- 1 expression construct into NIH 3T3 cells. In addition to the wild type, two mutant reporter constructs were also cotransfected with Gfi- 1. See A and B in Fig. 3C. In mutant A, the AATC core of both putative Gfi-1 binding sites was changed into ACTC, while in mutant B the AATC cores of the two sites were changed into AACT and AAGT, respectively. The mutant B base changes were designed so that the altered 18 bp repeats would match the repeat consensus. Based on the data presented in Fig. 2D1-3, Gfi-1 would not be expected to bind the

mutant sites. Cotransfection of Gfi-1 and the wild type HCMV-MIE promoter/CAT reporter construct downregulated the activity of the promoter more than four fold. Promoter mutations that abrogate Gfi-1 binding as shown in Fig. 2D1-3 also abrogated Gfi-1 mediated repression. See Fig. 3D.

REFERENCES

1. Horton, R. M., H. D. Hunt, S. N. So, J. K. Pullen, and L. R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene **77**:61-68.

2. Smith, D. B., and L. M. Corcoran. 1994. Expression and purification of glutathione S-transferase fusion proteins, section 16.7. In Ausubel (ed.), Current protocols in molecular biology. John Wiley & Sons, Inc.

3. Mavrothalassitis, G., G. Beal, and T. S. Papas. 1990. Defining target sequences of DNA-binding proteins by random selection and PCR: determination of the GCN4 binding sequence repertoire. DNA Cell Biol. **9**:783-788.

4. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd Ed., p. 1.25. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

5. Gogos, J. A., T. Hsu, J. Bolton, and F. C. Kafatos. 1992. Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. Science **257**:1951-1955.

6. Liu, C., W. S. Mason, and J. B. Burch. 1994. Identification of factor-binding sites in the duck hepatitis B virus enhancer and *in vivo* effects of enhancer mutations. J. Virol. **68**:2286-2296.

7. Baldwin, A. 1994. Methylation and uracil interference assays for analysis of proteinDNA

interactions, section 12.3. In Ausubel (ed.), Current protocols in molecular biology. John Wiley & Sons, Inc.

8. Glazman, Y. 1981. SV40-transformed simian cells support the replication of early SV4 mutants. Cell **23**: 175- 182.

9. Anderson, S., D. H. Davis, H. Dahlback, H. Jornvall, and D. W. Russell. 1989. Cloning, structure and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile biosynthetic enzyme. J. Biol. Chem. **264**:8222-8229.

10. Patriotis, C., A. Makris, J. Chernoff, and P. N. Tsichlis. 1994. Tpl-2 acts in concert with Ras and Raf-1 to activate the MAP kinase. Proc. Natl. Acad. Sci. USA **91**:9755-9759.

11. Andrews, N. C., and D. V. Faller. 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res. **19**:2499.

12. Thomsen, D. R., R. M. Stenberg, W. F. Goins, and M. F. Stinski. 1984. Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. Proc. Natl. Acad. Sci. USA **81**:659-663.

13. Ray, F. A., and J. A. Nickoloff. 1992. Site-specific mutagenesis of almost any plasmid using a PCR-based version of unique site elimination. Biotechniques **13**:342-346.

14. Bucher, P. 1995. The Eukaryotic Promoter Database EPD, EMBL Nucleotide Sequence Data Library, Release 43. European Bioinformatics Institute, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, United Kingdom.

15. Ghosh, D. 1993. Status of the transcription factors database (TFD). Nucleic Acids Res. **21**:3117-3118.

16. Dailey, L., S. M. Hanly, R. G. Roeder, and N. Heintz. 1986. Distinct transcription factors bind specifically to two regions of the human histone H4 promoter. Proc. Natl. Acad. Sci. USA **83**:7241-7245.

17. Crabtree, G. R. 1989. Contingent genetic regulatory events in T lymphocyte activation. Science **243**:355-361.

18. Ghazal, P., C. DeMattei, E. Ginlietti, S. A. Kliewer, K. Umesono, and R. M. Evans. 1992. Retinoic acid receptors initiate induction of the cytomegalovirus enhancer in embryonal cells. Proc. Natl. Acad. Sci. USA **89**:7630-7634.

19. Koedood, M., A. Fichtel, P. Meier, and P. J. Mitchell. 1995. Human cytomegalovirus (HCMV) immediate-early enhancer/promoter specificity during embryogenesis defines target tissues of congenital HCMV infection. J. Virol. **69**:2194-2207.

While certain preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made to the invention without departing from the scope and spirit thereof as set forth in the following claims.